PCT/GB 99/04445

The Patent Office

INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10

GB99/4445

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Signed Andrew Gerry

Dated 26 January 2000

Patents Form 1/77

Patents Act 1977
(6)

The Patent Office

04JAN99 E415322-1 000192
P01/7700 0.00 - 9828880.6

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form))

31 DEC 1998

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference — P.75768 GCW

2. Patent application number *(The Patent Office will fill in this part)* — **9828880.6** — 31 DEC 1998

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

ACTINOVA Limited
5 Signet Court
Swanns Road
Cambridge CB5 8LA

Patents ADP number *(if you know it)* 7415235001

If the applicant is a corporate body, give the country/state of its incorporation — UNITED KINGDOM

4. Title of the invention — PROTEIN

5. Name of your agent *(if you have one)* — J A KEMP & CO

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

14 SOUTH SQUARE
GRAY'S INN
LONDON WC1R 5LX

Patents ADP number *(if you know it)* 2660

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer "Yes" if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body:
See note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	33
Claim(s)	2
Abstract	1
Drawing(s)	6

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application

Signature J A Kemp & Co Date 31 December 1998

12. Name and daytime telephone number of person to contact in the United Kingdom

G.C. WOODS
0171 405 3292

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue of a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered "Yes" Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

PROTEIN

Field of the invention

This invention relates to a novel protein which forms part of an ABC transporter of *S.pyogenes*. The invention also relates to a streptococcus vaccine composition comprising this protein or fragments thereof.

Background of the invention

Streptococcus pyogenes, group A *Streptococcus*, is a common human pathogen which causes a variety of diseases such as pharingitis, impetigo, scarlatina and erysipelas. More severe infections caused by this organism are necrotizing fasciitis and streptococcal toxic shock like syndrome.

The superfamily of ABC (ATP-binding cassette) transporters comprise many different systems in procaryotes and eukaryotes. This diverse group of transporters serve many roles including transport of nutrients, translocation of signal molecules and chemotaxis. The general principle of ABC transport includes transportation of a ligand through two integral membrane domains forming a pore, with accompanying ATP hydrolysis by two nucleotide-binding domains associated with the cytoplasmic side of the pore.

In bacteria, the translocation of ligands is preceded by interaction with an accessory component, the periplasmic binding protein. This protein binds the ligand with higher affinity, and then interacts with the integral membrane components by releasing the ligand and allowing subsequent transport. In gram-positive bacteria, the binding protein homologue is a lipoprotein attached to the cell membrane by a NH_2-terminal lipid moiety. Little is known about the ABC transporters in gram-positive bacteria. In particular, the interaction between the lipid protein component and the integral membrane component is unclear.

A number of recent studies have looked at an ABC transporter family in *Streptococcus* species. Examples are Correia *et al* Infect. Immun (1996) 64(6) 2114-2121; Fenno *et al* Mol.Microbiol. (1995) 15(5) 849-863; Lowe *et al* Infect. Immun (1995) 63(2) 703-706; Kolenbrander *et al* Infect. Immun (1994) 62(10) 4469-4480

and Sampson *et al* Infect. Immun (1994) 62(1) 319-324.

The genes encoding the ABC transporter form an operon consisting of three genes. The putative proteins encoded by these genes are a hydrophobic membrane protein, a nucleotide binding protein and a lipid protein. The operon has been sequenced in a number of important disease causing organisms such as *S.pneumoniae*, *Enterococcus faecalis*, *S.sanguis* and *S.parasanguis* as well as in commensal bacteria such as *S.gordonii* and *Streptococcus crista*. As well as playing a role in transport, the ABC transporter has also been related to bacterial virulence and may mediate bacterial coaggregation, adhesion to host cells, saliva pellicle components and fibrin clots.

Summary of the invention

The applicants have now identified a new protein of *S.pyogenes* comprising the lipoprotein of an ABC transporter. The *S.pyogenes* operon is atypically organised and the polycistronic transcription is attenuated, in contrast to the previously described systems. The lipoprotein can be solubilized from the bacterial surface by proteolytic cleavage which indicates the presence of a flexible hinge region between the NH_2-terminal lipid moiety and a more compact globular fold. This lipoprotein and fragments thereof can be used in streptococcal vaccine compositions and in particular against *S.pyogenes*.

In a first aspect, the invention provides a polypeptide which comprises:

(a) the amino acid sequence of SEQ ID NO 1,

(b) a variant of (a) which is capable of binding an anti-SmtA antibody, or

(c) a fragment of (a) or (b) of at least 6 amino acids in length which is capable of binding to an anti-SmtA antibody.

In another aspect, the invention relates to a vaccine composition comprising a polypeptide which comprises:

(a) the amino acid sequence of SEQ ID NO 1,

(b) a variant of (a) which is capable of generating an immune response to a Streptococcus, or

(c) a fragment of (a) or (b) of at least 6 amino acids in length which is capable

of generating an immune response against a Streptococcus.

In a further aspect, the invention relates to novel polynucleotides having a sequence which is:

(i) the nucleotide coding sequence of SEQ ID NO 1 or a sequence complementary thereto,

(ii) a nucleotide sequence which selectively hybridises to a said sequence (i) or fragment thereof, or

(iii) a nucleotide sequence which codes for a polypeptide having the same amino acid sequence as that encoded by a said sequence of (i) or (ii).

Polynucleotides are therefore provided which encode a polypeptide of the invention. The invention also provides:

- a recombinant vector comprising a polynucleotide of the invention, such as an expression vector in which the polynucleotide is operably linked to a regulatory sequence;

- a host cell which is transformed with a polynucleotide of the invention;

- a process of producing a polypeptide of the invention comprising maintaining a host cell transformed with a polynucleotide of the invention under conditions to provide expression of the polypeptide;

- an antibody, monoclonal or polyclonal, specific for a polypeptide as defined in claim 1; and

- a method of vaccinating a patient against a Streptococcal infection, which method comprises administering to the patient an effective amount of a polypeptide according to the invention.

Description of the figures

Fig 1. **Organization of the *S.pyogenes lraI* operon.** The three genes in the *S.pyogenes lraI* operon are, from the left (5'), *smtA* (lipoprotein), *smtB* (ATP-binding protein), *smtC* (integral membrane protein). Arrows indicate the location of primers used to verify the operon organization, and to create probes for each of

the three genes. In the 66 bp non-coding region between *smtA* and *smtB* a putative stem-loop structure is present.

Fig 2. **Transcriptional analysis of the *S.pyogenes lraI* operon.** Total RNA (5 μg) from *S.pyogenes* bacteria in mid-log (ML) or early stationary (ES) growth phase was subjected to denaturing agarose gel electrophoresis and Northern blot hybridization. Three identical membranes were probed with radiolabelled PCR fragments corresponding to *smtA* (left panel), *smtB* (middle panel) or *smtC* (right panel).

Fig 3. **Surface localization of the SmtA protein on *S.pyogenes* bacteria.** A. Bacteria were washed and subjected to proteolytic digestion with papain at varying concentrations. Peptides solubilized were separated by SDS-PAGE (12%), and stained with Coomassie Brilliant Blue (stain). The peptides of an identical gel were transferred to a PVDF membrane and immunodetection with an antiserum directed against a synthetic SmtA peptide identified a band with an approximate molecular weight of 36 kDa (blot). B. The 36 kDa band was subjected to NH_2-terminal sequencing, and the sequence obtained was compared to the sequences of recombinant SmtA and database SmtA.

Fig 4. **Analysis of recombinantly expressed SmtA.** Fusion protein GST:SmtA and GST, purified on a glutathione Sepharose affinity column, and SmtA, purified by factor Xa cleavage of the fusion protein on the Sepharose matrix, were separated by 12% SDS-PAGE, and stained with Coomassie Brilliant Blue.

Fig 5. **Proton-induced X-ray emission analysis (PIXE).** PIXE analysis of GST, GST:SmtA, SmtA and (iron-saturated) transferrin in millipore water. Results are based on two different experiments and shown in mol element per mol protein.

Fig 6. **Binding of ^{65}Zn to SmtA.** A. SmtA and GST were separately incubated with ^{65}Zn. The mixtures were subjected to gel filtration on a PD-10 column and

500 μl fractions were collected. Fractions were assayed for protein content and radioactivity. Co-migration of ^{65}Zn and protein was seen in the SmtA sample (top graph) but not in the GST samples (bottom graph). B. GST or SmtA was applied in dilution series onto nitrocellulose membranes. The membranes were separately incubated with ^{65}Zn in citrate buffer (Na^+ concentration 0.1 M), pH 6.2, or with this buffer containing competing metal salts: 1 mM $ZnCl_2$, Fe(III)-citrate, Fe(II)-sulfate, $MnCl_2$, or $CuCl_2$, respectively. Computer analysis of the degree of inhibition was performed by comparing the radioactivity in the slits with the control incubation (top panel) lacking a competing salt.

Description of the sequences

SEQ ID NO 1 sets out the amino acid sequence for the full length SmtA polypeptide of *S.pyogenes* and the nucleotide sequence encoding this protein. The structure of SmtA is discussed in more detail below.

SEQ ID NO 2 sets out the amino acid sequence for full length SmtA of *S.pyogenes*.

Detailed description of the invention

The invention provides a polypeptide consisting essentially of (a) the amino acid sequence of SEQ ID NO 1; (b) a variant of the amino acid sequence of SEQ ID NO 1; or (c) a fragment of at least 6 amino acids in length of (a) or (b). Typically the polypeptide is capable of binding an anti-SmtA antibody.

Antibody to SmtA can be raised against purified SmtA protein using protein purified directly from *S.pyogenes* expressing this protein as described in more detail below. Alternatively, protein can be generated recombinantly. Following purification of the protein, antibody can be raised in an animal such as a rabbit and purified to generate the desired antibody. The antibody can be monoclonal or polyclonal antibody. Preferably, the polypeptide of the invention is incorporated in a vaccine composition for immunisation against a Streptococcal infection. Preferably, the antibody is neutralising antibody. Preferably, a polypeptide of the invention generates anti-SmtA antibody when administered *in*

vivo and provides protection against subsequent Streptococcal infection. Preferably, a polypeptide of the invention provides protection against group A Streptococcus but may also be used to provide vaccines against other Streptococcus such as *S.pnuemoniae*.

Polypeptides for incorporation into the vaccine compositions in accordance with the invention can be identified by determining whether they bind to an antibody specific for SmtA. Alternatively, antibody to a candidate polypeptide can be generated by standard techniques, for example by injection of the polypeptide into an appropriate animal and subsequent collection and purification of antisera from such animals. Antibody which binds SmtA can then be identified by standard and competitive immunoassays. The antibody thus identified can then be injected into mice to determine if it protects against a lethal challenge with a Streptococcal strain. Alternatively, an otherwise lethal dose of a strain of Streptococcus is given in an animal model system in which the animals have been given the relevant polypeptide.

As noted above a variant polypeptide (b) is one which will bind with an anti-SmtA antibody. Alternatively, a variant for incorporation into a vaccine composition is one which can be used to generate an immune response to provide protection against a Streptococcal infection. Over the entire length of the amino acid sequence of SEQ ID NO 1, a variant will preferably be at least 80% homologous to that sequence based on amino acid identity. More preferably, the polypeptide is at least 85% or 90% and more preferably at least 95%, 97% or 99% homologous to the amino acid sequence of SEQ ID NO 1 over the entire sequence. There may be at least 80%, for example at least 85%, 90% or 95%, amino acid identity over a stretch of 40 or more, for example 60, 100 or 120 or more, contiguous amino acids ("hard homology").

Amino acid substitutions may be made to the amino acid sequence of SEQ ID NO 1, for example from 1, 2 or 3 to 10, 20 or 30 substitutions. Conservative substitutions may be made, for example, according to the following table. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
	Polar - charged	D E
		K R
AROMATIC		H F W Y

One or more amino acid residues of the amino acid sequence of SEQ ID NO 1 may alternatively or additionally be deleted. From 1, 2 or 3 to 10, 20 or 30 residues may be deleted, or more. Polypeptides of the invention also include fragments (c) of the above-mentioned sequences. Such fragments retain the ability to bind anti-SmtA antibody. Fragments may be at least from 10, 12, 15 or 20 to 60, 100 or 200 amino acids in length. Particularly preferred fragments comprise;

- the sequence from amino acid residue number 136 through to residue 152 of SEQ ID NO 1, having the sequence KQLIAKDPKNKETYEKN;
- the sequence commencing at position 204 through to 222 of SEQ ID NO 1, having the sequence EINTEEEGTPDQISSLIEK;
- the sequence commencing at position 234 of SEQ ID NO 1 through to position 249 having the sequence ESSVDRRPMETVSKDS;
- the sequence commencing at position 259 of SEQ ID NO 1 through to position 279, having the sequence TDSIAKKGKPGDSYYAMMKWN; and
- variants of these sequences.

One or more amino acids may be alternatively or additionally added to the polypeptides described above. An extension may be provided at the N-terminus or C-terminus of the amino acid sequence of SEQ ID NO 1 or polypeptide variant or fragment thereof. The or each extension may be quite short, for example from 1 to 10 amino acids in length. Alternatively, the extension may be longer. A carrier protein may be fused to an amino acid sequence according to the invention. A fusion protein incorporating the polypeptides described above can thus be

provided.

Polypeptides of the invention may be in a substantially isolated form. It will be understood that the polypeptide may be mixed with carriers or diluents which will not interfere with the intended purpose of the polypeptide and still be regarded as substantially isolated. A polypeptide of the invention may also be in a substantially purified form, in which case it will generally comprise the polypeptide in a preparation in which more than 90%, e.g. 95%, 98% or 99%, by weight of the polypeptide in the preparation is a polypeptide of the invention.

Polypeptides of the invention may be modified for example by the addition of histidine residues to assist their identification or purification or by the addition of a signal sequence to promote their secretion from a cell where the polypeptide does not naturally contain such a sequence.

A polypeptide of the invention above may be labelled with a revealing label. The revealing label may be any suitable label which allows the polypeptide to be detected. Suitable labels include radioisotopes, e.g. ^{125}I, ^{35}S, enzymes, antibodies, polynucleotides and linkers such as biotin. Labelled polypeptides of the invention may be used in diagnostic procedures such as immunoassays in order to determine the amount of a polypeptide of the invention in a sample.

Polypeptides or labelled polypeptides of the invention may be used in serological or cell mediated immune assays for the detection of immune reactivity to said polypeptides in animals and humans. Standard protocols can be used. The labelled polypeptide may be used to identify and/or isolate "accessory" proteins which are involved in binding between cell receptors and SmtA, by detecting the interaction of SmtA to such proteins.

A polypeptide or labelled polypeptide of the invention or fragment thereof may also be fixed to a solid phase, for example the surface of an immunoassay well or dipstick.

Such labelled and/or immobilized polypeptides may be packaged into kits in a suitable container optionally including additional suitable reagents, controls or instructions and the like. The kits may be used to identify SmtA inhibitors or activators.

Such polypeptides and kits may also be used in methods of detection of antibodies to the SmtA protein by immunoassay.

Immunoassay methods are well known in the art and will generally comprise:

(a) providing a polypeptide comprising an epitope bindable by an antibody against said protein;

(b) incubating a biological sample with said polypeptide under conditions which allow for the formation of an antibody-antigen complex; and

(c) determining whether antibody-antigen complex comprising said polypeptide is formed.

Polypeptides of the invention may be made by synthetic means or recombinantly, as described below.

The polypeptides of the invention may be introduced into a cell by *in situ* expression of the polypeptide from a recombinant expression vector. The expression vector optionally carries an inducible promoter to control the expression of the polypeptide.

Such cell culture systems in which polypeptides of the invention are expressed may be used in assay systems.

A polypeptide of the invention can be produced in large scale following purification by high pressure liquid chromatography (HPLC) or other techniques after recombinant expression as described below.

A polynucleotide of the invention typically is a contiguous sequence of nucleotides which is capable of hybridising selectively with the coding sequence of SEQ ID NO 1 (nucleotides 1 to 861) or to the sequence complementary to that coding sequence. Polynucleotides of the invention include variants of the coding sequence of SEQ ID NO 1 which encode the amino acid sequence of SEQ ID NO 1 and variants and fragments of that sequence which are recognized by antibody to SmtA.

A polynucleotide of the invention and the coding sequence of SEQ ID NO 1 can hybridize at a level significantly above background. Background hybridization may occur, for example, because of other cDNAs present in a

cDNA library. The signal level generated by the interaction between a polynucleotide of the invention and the coding sequence of SEQ ID NO 1 is typically at least 10 fold, preferably at least 100 fold, as intense as interactions between other polynucleotides and the coding sequence of SEQ ID NO 1. The intensity of interaction may be measured, for example, by radiolabelling the probe, e.g. with ^{32}P. Selective hybridization is typically achieved using conditions of medium to high stringency (for example 0.03M sodium chloride and 0.03M sodium citrate at from about 50°C to about 60°C).

A nucleotide sequence capable of selectively hybridizing to the DNA coding sequence of SEQ ID NO 1 or to the sequence complementary to that coding sequence will be generally at least 80%, preferably at least 90% and more preferably at least 95%, homologous to the coding sequence of SEQ ID NO 1 or its complement over a region of at least 20, preferably at least 30, for instance at least 40, 60 or 100 or more contiguous nucleotides or, indeed, over the full length of the coding sequence. Thus there may be at least 85%, at least 90% or at least 95% nucleotide identity over such regions.

Any combination of the above mentioned degrees of homology and minimum size may be used to define polynucleotides of the invention, with the more stringent combinations (i.e. higher homology over longer lengths) being preferred. Thus for example a polynucleotide which is at least 85% homologous over 25, preferably over 30, nucleotides forms one aspect of the invention, as does a polynucleotide which is at least 90% homologous over 40 nucleotides.

Polynucleotides of the invention may comprise DNA or RNA. They may also be polynucleotides which include within them synthetic or modified nucleotides. A number of different types of modification to polynucleotides are known in the art. These include methylphosphate and phosphorothioate backbones, addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the polynucleotides described herein may be modified by any method available in the art.

Polynucleotides of the invention may be used to produce a primer, e.g a

PCR primer, a primer for an alternative amplification reaction, a probe e.g. labelled with a revealing label by contential means using radioactive or non-radioactive labels, or the polynucleotides may be cloned into vectors. Such primers, probes and other fragments will be at least 15, preferably at least 20, for example at least 25, 30 or 40 nucleotides in length, and are also encompassed by the term polynucleotides of the invention as used herein.

Polynucleotides such as a DNA polynucleotide and primers according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques. The polynucleotides are typically provided in isolated and/or purified form,

In general, primers will be produced by synthetic means, involving a step wise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art.

Longer polynucleotides will generally be produced using recombinant means, for example using PCR (polymerase chain reaction) cloning techniques. This will involve making a pair of primers (e.g. of about 15-30 nucleotides) to a region of the *smtA* gene which it is desired to clone, bringing the primers into contact with DNA obtained from a bacterial cell, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme recognition sites so that the amplified DNA can be cloned into a suitable cloning vector.

Such techniques may be used to obtain all or part of the *smtA* gene sequence described herein. Genomic clones containing the *smtA* gene and its promoter region may also be obtained in an analogous manner, starting with genomic DNA from a bacterial cell.

Although in general the techniques mentioned herein are well known in the art, reference may be made in particular to Sambrook *et al*, 1989.

Polynucleotides or primers of the invention may carry a revealing label. Suitable labels include radioisotopes such as ^{32}P or ^{35}S, enzyme labels, or other protein labels such as biotin. Such labels may be added to polynucleotides or primers of the invention and may be detected using techniques known *per se*.

Polynucleotides or primers of the invention or fragments thereof, labelled or unlabelled, may be used by a person skilled in the art in nucleic acid-based tests for detecting or sequencing SmtA in a sample.

Such tests for detecting generally comprise bringing a sample containing DNA or RNA into contact with a probe comprising a polynucleotide or primer of the invention under hybridizing conditions and detecting any duplex formed between the probe and nucleic acid in the sample. Such detection may be achieved using techniques such as PCR or by immobilizing the probe on a solid support, removing nucleic acid in the sample which is not hybridized to the probe, and then detecting nucleic acid which has hybridized to the probe. Alternatively, the sample nucleic acid may be immobilized on a solid support, and the amount of probe bound to such a support can be detected.

The probes of the invention may conveniently be packaged in the form of a test kit in a suitable container. In such kits the probe may be bound to a solid support where the assay formats for which the kit is designed requires such binding. The kit may also contain suitable reagents for treating the sample to be probed, hybridizing the probe to nucleic acid in the sample, control reagents, instructions, and the like.

Polynucleotides of the invention can be incorporated into a recombinant replicable vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus in a further embodiment, the invention provides a method of making polynucleotides of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells are described below in connection with expression vectors.

Preferably, a polynucleotide of the invention in a vector is operably linked

to a control sequence which is capable of providing for the expression of the coding sequence by the host cell, i.e. the vector is an expression vector. Such expression vectors can be used to express the polypeptide of the invention.

The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

Such vectors may be transformed into a suitable host cell to provide for expression of a polypeptide of the invention. Thus, a polypeptide according to the invention can be obtained by cultivating a host cell transformed or transfected with an expression vector as described above under conditions to provide for expression of the polypeptide, and recovering the expressed polypeptide.

The vectors may be for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of the said polynucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid. Promoters and other expression regulation signals may be selected to be compatible with the host cell for which the expression vector is designed.

Host cells transformed (or transfected) with the polynucleotides or vectors for the replication and expression of polynucleotides of the invention will be chosen to be compatible with the said vector and preferably will be bacterial such as *E.coli*. Alternatively they may be cells of a human or animal cell line such as CHO or COS cells, or yeast or insect cells. The cells may also be cells of a non-human animal such as a sheep or rabbit or plant cells.

The polypeptides of the invention are useful for vaccinating against a Streptococcal infection, for example against group A Streptococcus or *S.pneumoniae*. A vaccine of the invention comprises a suitable polypeptide and a pharmaceutically acceptable carrier or diluent. The preparation of vaccines which contain an immunogenic polypeptide(s) as active ingredient(s), is known to one

skilled in the art. Typically, such vaccines are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine.

Examples of adjuvants which may be effective include but are not limited to: aluminium hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamin (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutamnyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL + TDM + CWS) in a 2% squalene/Tween 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against an immunogenic polypeptide containing R28 antigenic sequence resulting from administration of this polypeptide in vaccines which are also comprised of the various adjuvants.

The vaccines are conventionally administered parentally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1% to 2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccarine, cellulose, magnesium carbonate, and

the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10% to 95% of active ingredient, preferably 25% to 70%. Where the vaccine composition is lyophilised, the lyophilised material may be reconstituted prior to administration, e.g. as a suspension. Reconstitution is preferably effected in buffer.

Capsules, tablets and pills for oral administration to a patient may be provided with an enteric coating comprising, for example, Eudragit "S", Eudragit "L", cellulose acetate, cellulose acetate phthalate or hydroxypropylmethyl cellulose.

The polypeptides of the invention may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric and maleic. Salts formed with the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine and procaine.

The vaccines are administrated in a manner compatible with the dosage formulation and in such amount as will be prophylactically effective. The quantity to be administered, which is generally in the range of 1 to 1000mg, preferably 5mg to 250mg, of polypeptide per dose, depends on a number of factors. These include the subject to be treated, capacity of the subject's immune system to synthesize antibodies, and the degree of protection desired. Precise amounts of active ingredient required to be administered may depend on the judgement of the practitioner and may be peculiar to each subject.

The vaccine may be given in a singe dose schedule, or preferably in a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals required to maintain and or reinforce the immune response, for example at 1 to 4 months for a second dose, and if needed, a

subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the need of the individual and be dependent upon the judgement of the practitioner.

The following Examples illustrate the invention.

Examples

Experimental procedures used in the following examples

Bacterial cultures

The *S.pyogenes* strains used in this study (serotypes M1, M4, M9, M12, M49) are from the World Health Organization Centre for references and research on Streptococci, Institute of Hygiene and Epidemiology, Prague, Czech Republic. The *S.pyogenes* strain SF370 is being sequenced in the Streptococcal Genome Project, and can be obtained from ATCC (700294). Streptococci were grown in Todd-Hewitt broth (Difco, Detroit, MI), supplemented with 0.2% yeast extract (Difco) in 5% CO_2 at 37°C. *E.coli* strain BL21 (Pharmacia Biotech, Uppsala, Sweden) was grown in Luria-Bertani or 2X YT broth or agar, aerobically at 37°C and supplemented with 100 μg/ml amplicillin (Sigma, St Louis, MO) or glucose where appropriate.

PCR, cloning procedures, and sequencing

Chromosomal DNA from *S.pyogenes* strains was extracted as described in Pitcher *et al*, Lett. App. Microbiol 1989 8 p151-156 modified in the initial incubation step by addition of 1000 U/ml of mutanolysin (Sigma) and 100 mg/ml lysozyme (Sigma). Oligonucleotide primers were designed by using sequence information from the Streptococcal Genome Project database, together with published sequences from other *Streptococcus* spp. Primers 5'-TAG-TAG-CGA-ATT-CGT-CGA-CTG-GCG-CTA-3' and 5'-AGC-ACA-ACT-CGA-GAA-TCG-CTG-TGC-TTT-A-3' enclose almost the whole of *smtA* (excluding the signal peptide and the NH_2-terminal cysteine residue) and were designed with an *EcoR*I

and *Xho*I restriction site, respectively. Primers 5'-GAT-TAC-AAC-TAA-CAA-TCT-TTG-TGT-GAC-C-3', 5'-TTG-ACA-AGG-TAT-CAA-CAG-TAA-ATA-CCT-C-3', 5'-ATG-TCA/T-CTC/T-ATG-GGA/G/T-GAT-GCC-ATC-3', and 5'-TTA/G-GCA-TAT/G-AG/AA-TAA/G-GCC/T-GTC-GC-3' were designed from internal segments of the genes *smtB* and *smtC*. PCR experiments were performed using *Taq* polymerase (Gibco-BRL, Gaithersburg, MD), except for cloning purposes, when *TaqPlus Precision*™ (Strategene, La Jolla, CA) was used. The PCT product corresponding to *smtA* was gel-purified prior to cloning, using Sephaglas™ Bandprep Kit (Pharmacia Biotech). The PCR profile consisted of 30 cycles at 94°C for 1 min, 50°C for 1 min, and 72°C for 2 min, followed by a final extension at 72°C for 7 min. Plasmid purification, restriction enzyme digestions, ligation, electroporation and screening of transformants were all performed according to standard procedures, or when applicable, according to instructions in the GST fusion protein kit (Pharmacia). Sequencing of the cloned insert was performed on an ABI-470 Prism with dyed dideoxy terminators, at Innovagen AB, Lund, Sweden.

Overexpression and purification of recombinant SmtA

An *E.coli* strain carrying the recombinant plasmid pGEX-5X-3:*smtA* was grown at 37°C overnight. This preculture was then inoculated 1:100 in prewarmed 2 x YT containing 100 μg/ml ampicillin. The culture was grown at 30°C until OD_{600}=1.2. Induction was then started by adding 0.5 mM IPTG (Promega, Madison, WI). After an additional 4 h of incubation the bacteria were pelleted by centrifugation at 8000 x g for 10 min, washed, and resuspended in PBS buffer, and lysed by sonication with a Branson sonifier B15 (Heimstatt, Germany). Triton X-100 was added to a final concentration of 1%, and the sample was gently mixed for 30 min. Cell debris was pelleted by centrifugation at 12000 x g for 10 min, and the supernatant was applied to a pre-equilibrated glutathione-sepharose suspension. All washes were performed in batch format, by centrifugation at 1000 x g for 5 min. GST:SmtA fusion protein was eluted with reduced glutathione. Alternatively, factor Xa cleavage (Pharmacia Biotech) was performed in 50 mM

Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM $CaCl_2$, overnight (1 U factor Xa/ml bacterial sonicate supernatant). Cleaved SmtA protein was then eluted in repeated steps with PBS.

RNA methods

Total RNA from *S.pyogenes* was purified using Fastprep™ cell disrupter (Savant, Holbrook, NY). Briefly, bacteria were cultured in THY medium until mid-logarithmic or early stationary phase, harvested by centrifugation at 3,800 x g for 10 min at 4°C, and resuspended in PBS, followed by disruption for 2 x 30 seconds at setting 6.0 using FastRNA™ kit with glass beads (BIO 101, Vista, CA) according to the manufacturers instructions.

For Northern blot experiments, RNA was separated on 1% agarose in 1 x HEPES buffer, blotted onto Hybond-N filters (Amersham, Amersham, UK), and hybridized with 600-900 bp long DNA probes specific for *smtA*, *smtB* and *smtC* (see above). The PCR products (see above) were purified on a MicroSpin™ S-200 HR column (Pharmacia, Uppsala, Sweden), and radiolabelled with [α-^{32}P]dATP using the Megaprime™ kit (Amersham). To verify that equal amounts of RNA were present on the filter, it was also probed with a radiolabelled 800 bp probe specific for 16S rRNA.

Protein methods

Protein samples were separated by 12% SDS-PAGE. Gels were then soaked in blotting buffer (20% ethanol, 200 mM glycine, 25 mM Tris), and proteins were blotted to an Immunobilon-P™ PVDF-membrane (Millipore). Membranes were then blocked in 15 ml of PBS, 0.05% Tween-20, 5% (w/v) skim milk for 20 min at room temperature. The primary antibody (see below) was added, diluted 1:1000 in PBS, 0.05% Tween-20 (PBST), and the membrane was incubated at 37 °C for 30 mins. Membranes were then washed in PBST, 3 x 5 min at 37°C. Horseradish peroxidase-conjugated antirabbit goat antibodies (Bio-Rad, Bio-Rad laboratories CA) diluted 1:3000 in PBST were added and the membrane was incubated for 30 min at 37°C, and then washed as above.

Proteolytic digestion of bacteria was performed on overnight cultures of *S. pyogenes*. Bacteria were pelleted by centrifugation at 3000 x g for 5 min, at 8°C, resuspended in cleavage buffer (0.01 M Tris-HCl, pH 8), and washed twice as above before being resuspended in cleavage buffer. Papain (Sigma) was added (0 - 200 $\mu g/2 \times 10^9$ bacteria), and digestion was initiated by the addition of L-cysteine (55 mM). The suspension was incubated at 37°C for 1 h. The digestion was stopped by adding Iodoacetamide to a final concentration of 12 μM. Bacteria were then pelleted by centrifugation at 4800 x g for 10 min, and the supernatant was removed and filtered through a 0.2 μm Acrodisc (Gelman Sciences, Ann Arbor, MI). After freezing and thawing, samples precipitated spontaneously. This was used as a convenient method of concentration, as SDS-PAGE analysis indicated that the precipitate contained all the proteins present in the original solubilized sample. Enzymatic digestion with streptococcal cysteine proteinase was performed essentially as described in Berge and Björck, J. Biol Chem (1995) 270 9862-9867.

Metal Assays

Samples for PIXE analysis were prepared as follows. One ml of GST, SmtA fusion, and SmtA suspended in PBS was ultrafiltrated in Centricon-10 cells (Amicon, Inc., Beverly, MA) four times, each time reducing the volume to less than one tenth, and adding Millipore water up to 1.5 ml after the first three spins. The final protein concentration was determined using Coomassie® Protein Assay Reagent kit (Pierce, Rockford, IL). Bovine iron-saturated holo-Transferrin (Sigma) was suspended in Millipore water. Then, 20 and 60 μg of GST, SmtA fusion and SmtA, respectively, in 25 μl of Millipore water was added to a Kimfoil (Kimberley Clark), mounted on a plastic frame, and allowed to dry. 200 μg of transferrin in 25 μl of Millipore water was similarly prepared. At the Lund Nuclear Microprobe the samples were placed in the vacuum irradiation chamber and bombarded with a 2 nA proton beam, having an energy 2.55 MeV, accumulating a beam charge of 0.6 μC. The characteristic X-rays emitted were detected using a Kevex 50 mm^2 Si(Li) X-ray detector, while data were collected using a CAMAC/Mac-computer system

equipped with KMAX (Sparrow) software. Elemental standards (Fe, Co, Ni) were analysed in the same batch to verify quantification.

Gel filtration experiments were performed with PD-10 columns (Pharmacia Biotech). 20 μg protein (GST or SmtA) in PBS, supplemented with 0.25 M NaCl, was incubated overnight at 4°C with approx. 0.5 μCi of ^{54}Mn, ^{65}Zn or ^{59}Fe (Amersham). The sample was applied to a gel filtration column, preequilibrated with PBS, 0.25 M NaCl, and 0.5 ml fractions were eluted with the same buffer. Each fraction was then assayed for protein content, using Coomassie® Protein Assay Reagent kit (Pierce), and radioactivity, by adding Ready Safe™ scintillation fluid (Beckman Instruments, Fullerton, CA) and counting β-emission on a Beckman Instruments LS6000TA.

Slot blot experiments were performed by applying 10, 1 and 0.1 μg of protein (SmtA fusion and GST) to a nitrocellulose membrane, using a slot blot apparatus (Millipore). Membranes were equilibrated in 0.1 M citrate buffer, pH 6.2, (appr. 0.1 M Na^+), or 0.1 M maleic acid, pH 6.5, 0.1 M NaCl (for iron-binding) and then suspended in the same buffer with 1 μCi/ml of radioisotope (^{59}Fe, ^{54}Mn or ^{65}Zn), for 1 h at room temperature and with gentle mixing. In competition experiments 1 mM of Fe(II) sulfate, Fe(III) citrate, $MnCl_2$, $CuCl_2$ or $ZnCl_2$ was added. Then, membranes were washed for 2 x 15 min in buffer, and exposed on a phosphoimaging plate (Fuji Photo Film Co. Ltd., Japan). Quantification of bound radioactivity was performed using the Bio-Imaging Analyzer BAS2000 program package (Fuji Photo Film).

Other Methods

NH_2-terminal sequencing (Edman degradation) was performed at the Biomedical Service Unit, Lund University. The peptide QDPHEYEPLPEDV was synthesized, analysed for purity and correct sequence.

Rabbits were immunized with peptide-KLH conjugates. Primary immunization was performed with 100 μg peptide-KLH conjugate in Freunds complete adjuvant (Sigma). The two booster immunizations were done at weeks 4 and 6, using 100 μg peptide-KLH conjugate in Freunds incomplete adjuvant

(Sigma).

Example 1 - *Identification and Sequence analysis of a S. pyogenes member of the lraI family*

Predicted amino acid sequences of the proteins encoded by the various *lraI* operons were used to search (tBLASTn) the Streptococcal Genome Project database, at the time of 95% completion. The products of four adjacent open reading frames (ORF), showed strong homology to the LraI proteins. A frame shift split one gene (lipoprotein) in two ORF's but subsequent sequencing (see below) of a serotype M1 strain showed that the database sequence was incorrect, containing a single base insertion in a region of reduced database sequence quality. The (corrected) three ORFs were named *smtABC* (streptococcal metal transporter).

The three ORFs encode proteins typical of an ABC transport system. SmtA contains a putative hydrophobic signal peptide and a consensus sequence (*LXXC*) typical of NH_2-terminal lipid linkage bacterial lipoproteins. SmtB has an ATP-binding cassette, while SmtC is a highly hydrophobic protein with 6-7 potential transmembrane domains. Homologies between the putative proteins of *S. pyogenes* and their counterparts in other *Streptococcus* spp. were in the same range as previously described for the *lraI* family, as noted in the Table below. The percentages indicate sequence identity at the protein level. An "*" indicates only partial sequences available.

Bacterium	Lipoprotein	ATP-binding protein	Integral membrane protein
S.pneumoniae	72%	51%*	70%
S.gordonii	73%	56%	78%
S.parasanguis	72%	54%	75%
S.sanguis	74%		75%*
S.crista	72%		73%*
E.faecalis	55%		

Notably, the arrangement of the genes in the *S. pyogenes* operon was atypical. The lipoprotein gene is at the 5'-end of the operon, whereas in all other operons it is at the 3'-end. A publication on an oligopeptide permease ABC-transporter in *S. pyogenes* Polbielski *et al* Mol. Microbiol. 1996 21(5) 1087-1099 showed a differential transcription of the lipoprotein component. A putative stem-loop structure (-60.2 kJ mol^{-1}) (Fig. 1B) similar to Rho-independent transcription terminators was identified in the 66 bp non-coding region between *SmtA* and *SmtBC*.

The atypical organization was verified by PCR experiments. Forward and reverse primer pairs from each of the three genes were used to amplify products with template DNA from five different strains (serotypes M1, M4, M9, M12, and M49) of *S. pyogenes*, in combinations so as to allow every possible arrangement of the three genes in the operon. All experiments gave single products of a size consistent only with the atypical operon seen in the database (data not shown).

The GENBANK database was searched for homologues of SmtA, SmtB and SmtC, using the BLAST algorithm. The highest scoring matches were all ABC transporters specific for di-or trivalent cations. Specifically, several iron-siderophore transport systems showed a high degree of homology to the streptococcal proteins. SmtA has a 34% identify with YfeA (Bearden *et al* J. Bacteriol 1998 180(5) 1135-1147), the periplasmic binding protein of an iron (chelated) transport system in *Yersinia pestis* (also in *Haemophilus influenzae*). An inversely directed search, using YfeA to search for homologues in the Streptococcal Genome database, identified SmtA as the best homologue on the streptococcal genome. TroA (Hardman *et al* Gene 1997 197 47-64) from *Treponema pallidum* also shows homology (27% id. and 17% sim.) with SmtA, and is part of an ABC transporter operon flanked by iron-regulated transcription factors. Additionally, a recently described iron-regulated ABC transporter in *Staphylococcus epidermidis* (Cockayne *et al* Infect Immun 1998 66 3767-3774) has a high degree of homology (the *S. epidermidis* lipoprotein is 52% identical to SmtA) to IraI protein, and should possibly be included in this family.

Two homologous ABC transporters not involving iron were also found:

the manganese transporter MntABC (Bartsevich and Pakrasi EMBO J 1995 14(9) 1845-1853) from *Synechococcus cystis*, and the zinc transporter AdcCBA from *S. pneumoniae*, where SmtA showed appr. 30% identity with the corresponding proteins, MntC and AdcA. A comparison of SmtC with conserved motifs of integral membrane proteins from ABC transporters (Saurin *et al* Mol Microbiol 1994 12 993-1004) showed that the best fitting (21% id. and 21% sim. on 43 aa) motif was from the cluster of iron-siderophore transporters.

Example 2 - *Transcriptional analysis of smtABC*

In order to investigate the transcription of the three genes in the *S. pyogenes lraI* operon, probes corresponding to internal sequences of *smtA*, *smtB* and *smtC* were produced and labelled. Total RNA was extracted from serotype M1 *S. pyogenes* bacteria grown to mid-logarithmic and early stationary phase. Northern blot experiments (Fig. 2) showed that all three probes reacted weakly with a transcript approximately 2.5 kb in size when RNA from mid-log phase bacteria was used, consistent with a polycistronic transcription of the operon. However, the *smtA* probe also reacted with a shorter transcript, approximately 1 kb in size. This shorter transcript was present in higher (10 -20 times) amounts than the polycistronic transcript, suggesting that the stem-loop structure in the non-coding region between *smtA* and *smtB* can terminate transcription. The *SmtA* probe showed weak reactivity with the short transcript when RNA from bacteria in early stationary phase was used, whereas the two other probes did not hybridize at all. A control hybridization with a probe for 16S rRNA was done to verify that RNA levels in the samples (mid-log and early stationary phase of growth) were equal (data not shown). Taken together, these results indicate that the lipoprotein is expressed in higher quantities than the ATPase and hydrophobic membrane protein.

Example 3 - *Surface localization of the SmtA lipoprotein*

We produced the synthetic peptide QDPHEYEPLPEDV, spanning a highly conserved region of SmtA, predicted to have good antigenicity. A rabbit

was immunized with the peptide, and following three boosters the serum showed good reactivity with the peptide in ELISA, as compared with preimmune serum (data not shown). Then, a proteolytic digestion of *S. pyogenes* bacteria was performed, to investigate whether a protein fragment from the lipoprotein could be identified using the peptide antiserum. The protein fragments released by proteolytic digestion were subjected to SDS-PAGE and then transferred to PVDF-membrane by Western blotting.

Immunodetection with the peptide antiserum identified a protein fragment with an apparent molecular mass of 36 kDa, solubilized at high concentrations of papain (Fig 3A). The protein seemed fairly resistant to proteolytic digestion, since it remained at the same position even at high papain concentrations when most of the proteins had shifted to the low molecular weight range, supposedly degraded by the excess of protease. Protein from the 36 kDa band was subjected to NH_2-terminal amino acid sequencing. The result (Fig 3B) showed a 90% identity to a region very close to the NH_2-terminus in the predicted SmtA protein (amino acids 30-39) of the database, and a 100% identity to the predicted SmtA protein from the sequence of the strain studied (see below). The predicted molecular mass of the mature SmtA polypeptide was 32 kDa. Thus, papain seems to cleave SmtA very close to the protein's predicted NH_2-terminal lipid anchor, liberating almost the whole polypeptide from the bacterial cell surface. A similar papain digestion and western blot was also performed with the strain SF370 sequenced in the Streptococcal Genome Project, with the same result.

Bacterial growth media from overnight cultures of *S. pyogenes* was also examined for presence of SmtA, since bacterial lipoproteins are sometimes found in medium as well. No reactivity with the peptide antiserum was seen in TCA-precipitated proteins from medium. Also, mild detergent treatment of bacteria failed to solubilize any protein reacting with the antiserum.

S. pyogenes produces and secretes a cysteine proteinase, SCP. This protease has previously been shown to release functionally active fragments of streptococcal surface proteins. We performed proteolytic digestions of *S. pyogenes* with purified SCP. Solubilized proteins were visualized by SDS-PAGE and analysed by

immunoblotting. There was no reactivity with the antiserum. To exclude that SmtA had been completely degraded, a double digestion was performed, first with SCP and then with papain. The amount of SmtA released, as judged by staining and immunoblotting, was unaffected by pretreatment with SCP (data not shown).

Example 4 - *Cloning and purification of recombinant SmtA*

The *smtA* gene from strain AP1 was PCR-amplified and cloned into pGEX and sequence analysis confirmed the presence of *SmtA*, showing 98% (amino acid) identity with the database sequence, well in line with what could be expected from two different strains of the same serotype (compare *psa*A from *S. pneumoniae* Berry and Paton Infect. Immun. 1996 64(12) 5255-5262. There was no frame shift in the sequenced ORF, indicating that the preliminary database sequence indeed contained an incorrect base insertion (see above). Also, the supposed error is found in a region where the database sequence is less accurate. SmtA was purified by affinity chromatography of over expressed SmtA fusion and subsequent proteolytic cleavage with factor Xa. The purified SmtA had an apparent molecular mass of 36 kDa, and comprised > 95% of the protein content in the sample (Fig. 4).

Computer predictions (Robson-Garnier and Chou-Fasman algorithms) of SmtA secondary structure suggest a predominantly α-helical structure (45-60% of the protein). CD-spectroscopy of SmtA confirmed these predictions (data not shown).

Example 5 - *Analysis of trace element content and metal binding properties of SmtA*

A recent publication on a periplasmic molybdate-binding protein of *E. coli* Rech *et al* J. Biol Chem 1996 271 2557-2562 showed that binding of the ligand changed the migration of the protein in native PAGE. Similar experiments were performed with SmtA, but no consistent effects on migration properties was seen with any of the tested potential ligands (Mn(II), Fe(III), Cu(II), Zn(II), Co(II), Ca(II)), nor with EDTA (data not shown). We performed a quantitative analysis of trace element content in SmtA, SmtA fusion, GST and (iron-saturated)

transferrin by the use of a highly specific and sensitive technique, proton-induced X-ray emission (PIXE). PIXE analysis indicated that iron was present in approximately 1.5 molar ratio compared to protein in the SmtA sample (Fig. 5). Little or no iron was detected in the other samples, except for transferrin. Among other trace elements only copper was present in significant amounts. Both SmtA fusion and SmtA contained copper in approximately 0.5 molar ratio to protein. In a similar analysis, using a 0.1 M Tris-Ac buffer, pH 7.5, a significant Zn content was found (data not shown).

In addition, 30 μM solutions of GST and SmtA were analysed for iron content at a routine clinical chemistry laboratory. The sample with SmtA contained 20 μM iron, indicating a 67% saturation of the protein. To rule out that iron contamination from the factor Xa solution used for cleaving the fusion protein was affecting the results, a sample containing cleavage solution, reduced glutathione, and PBS was analysed for iron content. Iron concentration in that sample was <5 μM (detection limit).

SmtA was subjected to proteolytic digestion with trypsin, papain, and proteinase K (5:1 ratio SmtA/protease). Digestion patterns of SmtA with/without pretreatment with EDTA were compared, to elucidate whether a potentially present cation ligand affected the conformation of the protein so as to change accessibility of proteolytic cleavage sites. No such effect was seen (data not shown). A relative resistance to proteolytic digestion with papain was noted, in accordance with the case of the native protein (see above). In addition a high degree of trypsin resistance was noted.

We chose to investigate direct binding of metal radioisotopes (^{59}Fe, ^{65}Zn and ^{54}Mn) to the recombinant protein. Iron proved to be notoriously difficult to work with, since ferric iron is essentially insoluble at neutral pH. Precipitation problems therefore prevented reasonable interpretation of several assays tried. Copper radioisotopes decay too rapidly to handle practically. Interestingly, an interaction with Zn could be demonstrated. GST and SmtA were incubated with ^{65}Zn, in the presence of 0.25 M NaCl, and then subjected to gel filtration. Fractions were collected, and their protein concentration and radioactivity was

measured. The result (Fig. 6A) indicates comigration of ^{65}Zn with SmtA, but not with GST. When the molar amount of comigrating ^{65}Zn is compared with the protein content, a 60% saturation of the protein is found (assuming a single binding site). Similarly incubated GST and SmtA were subjected to native PAGE, and then autoradiography. Radioactivity was seen at the place of SmtA, but not for GST (data not shown), and this could be inhibited by adding $ZnCl_2$ to the initial incubation.

A slot blot assay was also performed, applying SmtA and GST onto a nitrocellulose membrane in dilution series. The membrane was then incubated with ^{65}Zn, and, following washing steps, radioactivity was found in the SmtA fusion sample, but not in the GST sample (Fig 6B). This interaction could be efficiently inhibited (88%) by addition of 1 mM $ZnCl_2$ during incubation (Fig 6B). A varying degree of inhibition was seen when adding other metal salts (Fig 6B). Among these, only Cu(II) inhibited the interaction to a high degree (74%). A similar experiment was performed with ^{59}Fe(III), which also bound to SmtA but not to GST (data not shown). The binding was completely inhibited by addition of 1 mM Fe(III) citrate, but only weakly inhibited by 1 mM $CuCl_2$ and $MnCl_2$. Addition of Fe(II)-sulfate or $ZnCl_2$ caused unspecific interaction with the membrane, probably due to precipitation. All these assays were also performed with ^{54}Mn, but no interaction between the radioisotope and any of the proteins was found.

SEQUENCE LISTING

<160> 2

<170> PatentIn Ver. 2.0

<210> 1

<211> 893

<212> DNA

<213> Streptococcus pyogenes

<220>

<221> CDS

<222> (1)..(861)

<400> 1

tcg tcg act ggc gct aaa act gct aag agt gat aaa tta aaa gtc gtg 48
Ser Ser Thr Gly Ala Lys Thr Ala Lys Ser Asp Lys Leu Lys Val Val
1 5 10 15

gca acc aat tca att att gcc gac atg aca aaa gct att gct ggt gat 96
Ala Thr Asn Ser Ile Ile Ala Asp Met Thr Lys Ala Ile Ala Gly Asp
20 25 30

```
aaa atc gat ctg cac agc att gtg cca atc ggt caa gac cct cat gag  144
Lys Ile Asp Leu His Ser Ile Val Pro Ile Gly Gln Asp Pro His Glu
        35                  40                  45

tac gaa cca tta cca gaa gat gct gaa aaa aca agt aat gct gat gtg  192
Tyr Glu Pro Leu Pro Glu Asp Ala Glu Lys Thr Ser Asn Ala Asp Val
    50                  55                  60

att ttc tat aat ggt atc aat cta gaa gat ggc ggg caa gct tgg ttc  240
Ile Phe Tyr Asn Gly Ile Asn Leu Glu Asp Gly Gly Gln Ala Trp Phe
65                  70                  75                  80

acc aaa cta gtg aaa aat gct caa aaa acg aaa aac aaa gat tac ttt  288
Thr Lys Leu Val Lys Asn Ala Gln Lys Thr Lys Asn Lys Asp Tyr Phe
                85                  90                  95

gcc gtg tct gat ggc att gat gtg att tac ttg gaa ggt gca agc gaa  336
Ala Val Ser Asp Gly Ile Asp Val Ile Tyr Leu Glu Gly Ala Ser Glu
            100                 105                 110

aaa gga aaa gaa gat cca cat gct tgg tta aat ctc gaa aac gga atc  384
Lys Gly Lys Glu Asp Pro His Ala Trp Leu Asn Leu Glu Asn Gly Ile
        115                 120                 125

att tat tca aaa aac att gcc aaa caa ttg att gca aag gat cct aaa  432
Ile Tyr Ser Lys Asn Ile Ala Lys Gln Leu Ile Ala Lys Asp Pro Lys
    130                 135                 140

aac aaa gaa act tat gaa aag aac cta aaa gct tat gtg gct aaa ttg  480
Asn Lys Glu Thr Tyr Glu Lys Asn Leu Lys Ala Tyr Val Ala Lys Leu
145                 150                 155                 160
```

gaa aaa cta gac aaa gaa gcc aaa tca aaa ttt gat gct att gca gaa 528
Glu Lys Leu Asp Lys Glu Ala Lys Ser Lys Phe Asp Ala Ile Ala Glu
165 170 175

aat aaa aaa ttg att gtg act agt gaa ggc tgc ttc aag tac ttt tca 576
Asn Lys Lys Leu Ile Val Thr Ser Glu Gly Cys Phe Lys Tyr Phe Ser
180 185 190

aaa gct tac ggt gtc cca tct gct tat atc tgg gaa att aac acc gaa 624
Lys Ala Tyr Gly Val Pro Ser Ala Tyr Ile Trp Glu Ile Asn Thr Glu
195 200 205

gaa gaa gga aca cca gat caa att tca tca ttg att gaa aaa cta aaa 672
Glu Glu Gly Thr Pro Asp Gln Ile Ser Ser Leu Ile Glu Lys Leu Lys
210 215 220

gtc atc aag cca tct gcg ctt ttt gta gag tca agt gtc gat aga cgc 720
Val Ile Lys Pro Ser Ala Leu Phe Val Glu Ser Ser Val Asp Arg Arg
225 230 235 240

cct atg gaa act gtt tct aaa gat agt ggt att cct att tat tct gag 768
Pro Met Glu Thr Val Ser Lys Asp Ser Gly Ile Pro Ile Tyr Ser Glu
245 250 255

atc ttt aca gat tca att gct aaa aaa ggt aaa cct ggc gat agt tat 816
Ile Phe Thr Asp Ser Ile Ala Lys Lys Gly Lys Pro Gly Asp Ser Tyr
260 265 270

tat gct atg atg aaa tgg aac ctt gac aaa att tct gaa ggt cta 861
Tyr Ala Met Met Lys Trp Asn Leu Asp Lys Ile Ser Glu Gly Leu
275 280 285

gcaaaataag ataacaataa agcacagcga tt 893

<210> 2

<211> 287

<212> PRT

<213> Streptococcus pyogenes

<400> 2

Ser Ser Thr Gly Ala Lys Thr Ala Lys Ser Asp Lys Leu Lys Val Val
1 5 10 15

Ala Thr Asn Ser Ile Ile Ala Asp Met Thr Lys Ala Ile Ala Gly Asp
20 25 30

Lys Ile Asp Leu His Ser Ile Val Pro Ile Gly Gln Asp Pro His Glu
35 40 45

Tyr Glu Pro Leu Pro Glu Asp Ala Glu Lys Thr Ser Asn Ala Asp Val
50 55 60

Ile Phe Tyr Asn Gly Ile Asn Leu Glu Asp Gly Gly Gln Ala Trp Phe
65 70 75 80

Thr Lys Leu Val Lys Asn Ala Gln Lys Thr Lys Asn Lys Asp Tyr Phe
85 90 95

Ala Val Ser Asp Gly Ile Asp Val Ile Tyr Leu Glu Gly Ala Ser Glu
100 105 110

Lys Gly Lys Glu Asp Pro His Ala Trp Leu Asn Leu Glu Asn Gly Ile
115 120 125

Ile Tyr Ser Lys Asn Ile Ala Lys Gln Leu Ile Ala Lys Asp Pro Lys
130 135 140

Asn Lys Glu Thr Tyr Glu Lys Asn Leu Lys Ala Tyr Val Ala Lys Leu
145 150 155 160

Glu Lys Leu Asp Lys Glu Ala Lys Ser Lys Phe Asp Ala Ile Ala Glu
165 170 175

Asn Lys Lys Leu Ile Val Thr Ser Glu Gly Cys Phe Lys Tyr Phe Ser
180 185 190

Lys Ala Tyr Gly Val Pro Ser Ala Tyr Ile Trp Glu Ile Asn Thr Glu
195 200 205

Glu Glu Gly Thr Pro Asp Gln Ile Ser Ser Leu Ile Glu Lys Leu Lys
210 215 220

CLAIMS

1. A polypeptide which comprises:
   (a) the amino acid sequence of SEQ ID NO 1,
   (b) a variant of (a) which is capable of binding an anti-SmtA antibody, or
   (c) a fragment of (a) or (b) of at least 6 amino acids in length which is capable of binding an anti-SmtA antibody.

2. A vaccine composition comprising a polypeptide which comprises:
   (a) the amino acid sequence of SEQ ID NO 1,
   (b) a variant of (a) which is capable of generating an immune response to a Streptococcus or
   (c) a fragment of (a) or (b) of at least 6 amino acids in length which is capable of generating an immune response against a Streptococcus.

3. A vaccine composition according to claim 2 wherein the Streptococcus is a group A Streptococcus.

4. A polynucleotide having a sequence which is
   (i) the nucleotide coding sequence of SEQ ID NO 1 or the sequence complementary thereto,
   (ii) a nucleotide sequence which selectively hybridises to a said sequence (i) or a fragment thereof, or
   (iii) a nucleotide sequence which codes for a polypeptide having the same amino acid sequence as that encoded by a said sequence (i) or (ii).

5. A polynucleotide according to claim 4 wherein the sequence (i), (ii) or (iii) encodes a polypeptide capable of generating an immune response to a Streptococcus.

6. An expression vector comprising a polynucleotide according to claim 4 or 5

operably linked to a regulatory sequence.

7. A host cell transformed with a polynucleotide of claim 4 or claim 5.

8. A process of producing a polypeptide suitable for use in vaccination against a Streptococcus comprising maintaining the host cell as defined in claim 7 under conditions to provide expression of the polypeptide and recovering the expressed polypeptide.

9. An antibody specific for a polypeptide as defined in claim 1.

10. A method of vaccinating a patient against a Streptococcal infection, which method comprises administering to the patient an effective amount of a polypeptide as defined in claim 1.

ABSTRACT

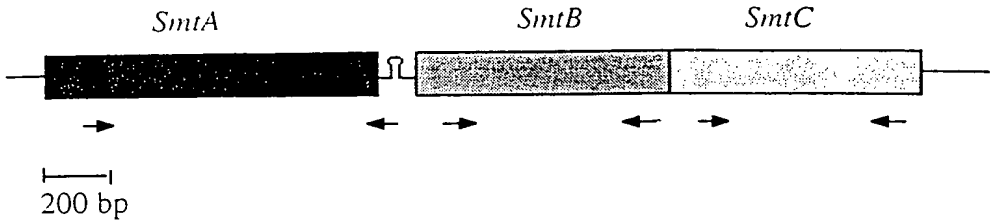
PROTEIN

A polypeptide suitable for use in vaccination against Streptococcal infections comprises:

(a) the amino acid sequence of SEQ ID NO 1.

(b) a variant of (a) which is capable of binding an anti-SmtA antibody, or

(c) a fragment of (a) or (b) of at least 6 amino acids in length which is capable of binding an anti-SmtA antibody.


SmtA
SmtB
SmtC
200 bp

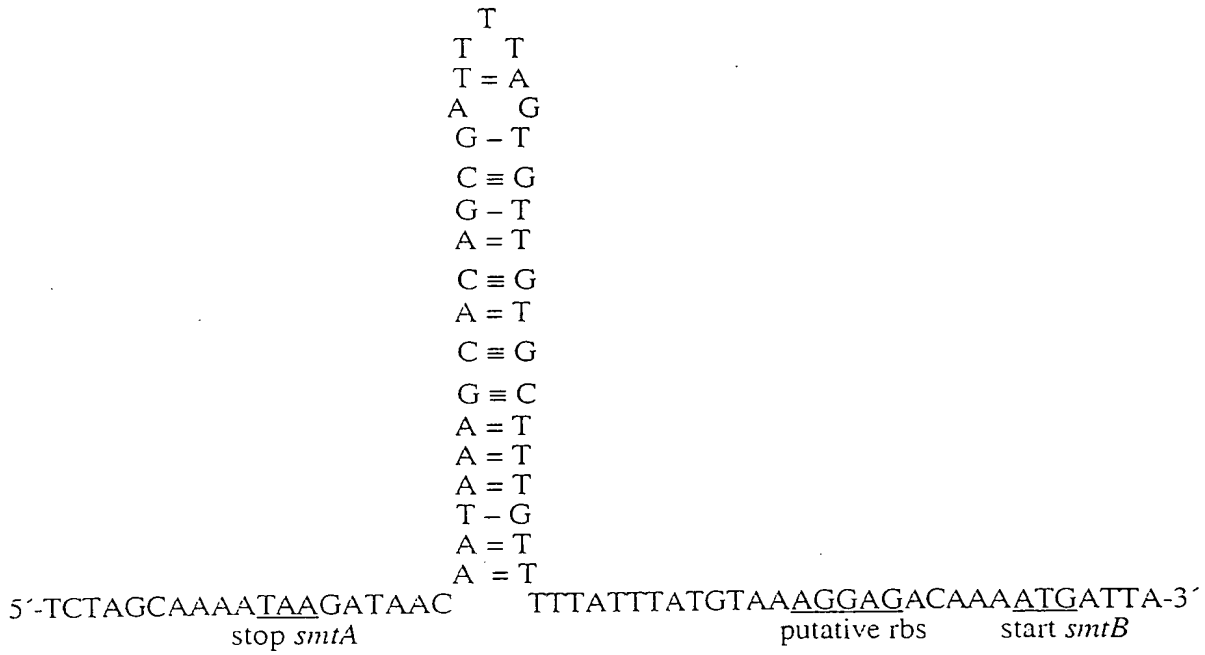

T
T T
T = A
A G
G – T
C ≡ G
G – T
A = T
C ≡ G
A = T
C ≡ G
G ≡ C
A = T
A = T
A = T
T – G
A = T
A = T
5′-TCTAGCAAAATAAGATAAC TTTATTTATGTAAAGGAGACAAAATGATTA-3′
stop smtA
putative rbs
start smtB

FIG. 1

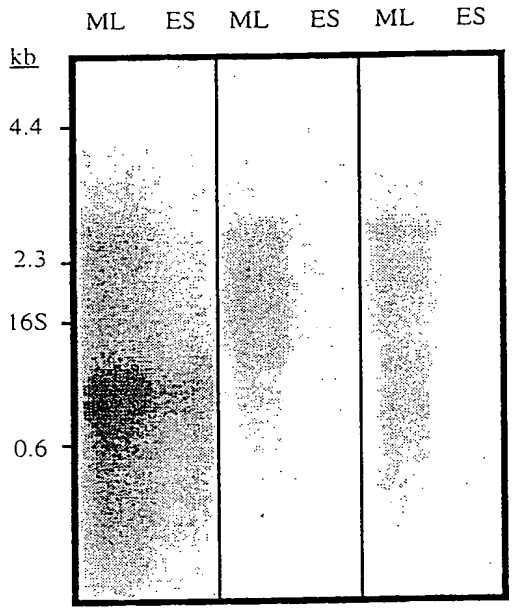
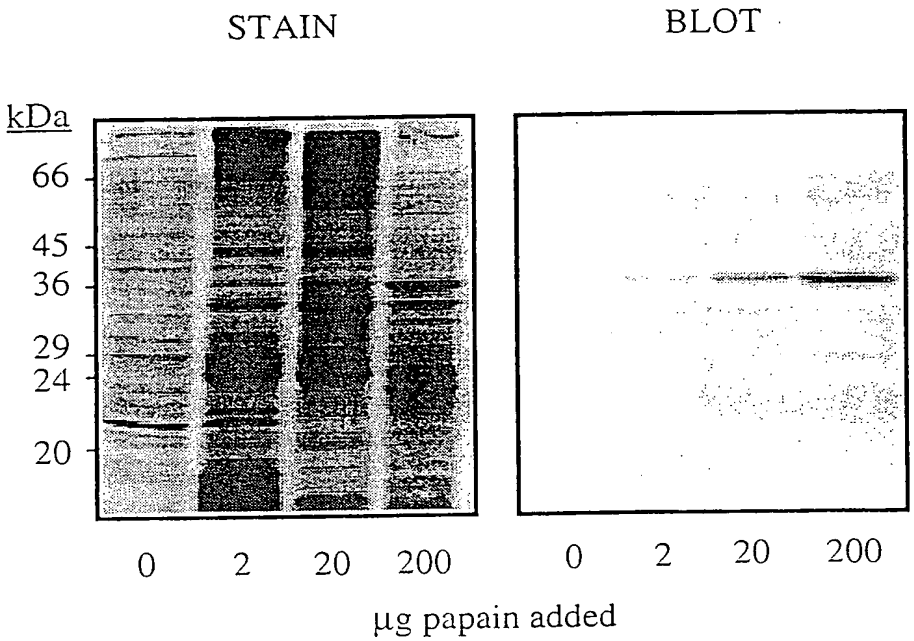


FIG. 2

A



B

36 kDa papain fragment (aa 1-10) **KSDKLKVVAT**

recombinant SmtA (aa 30-39) **KSDKLKVVAT**

database SmtA (aa 30-39) ***E*SDKLKVVAT**

FIG. 3

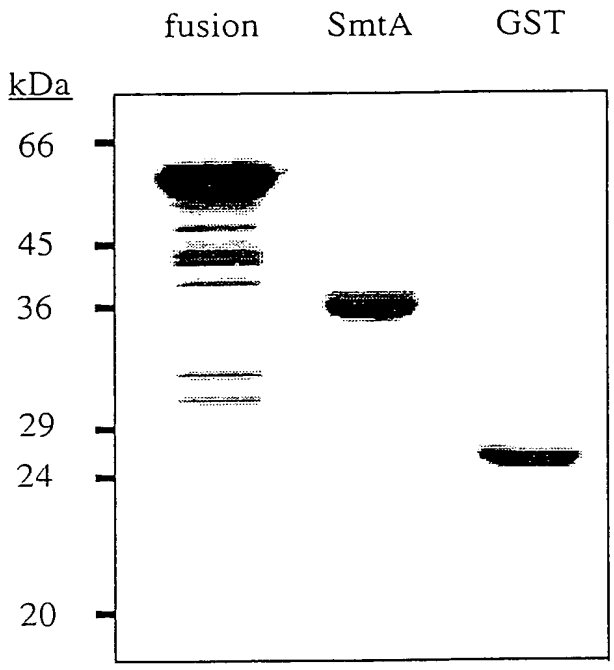


FIG. 4

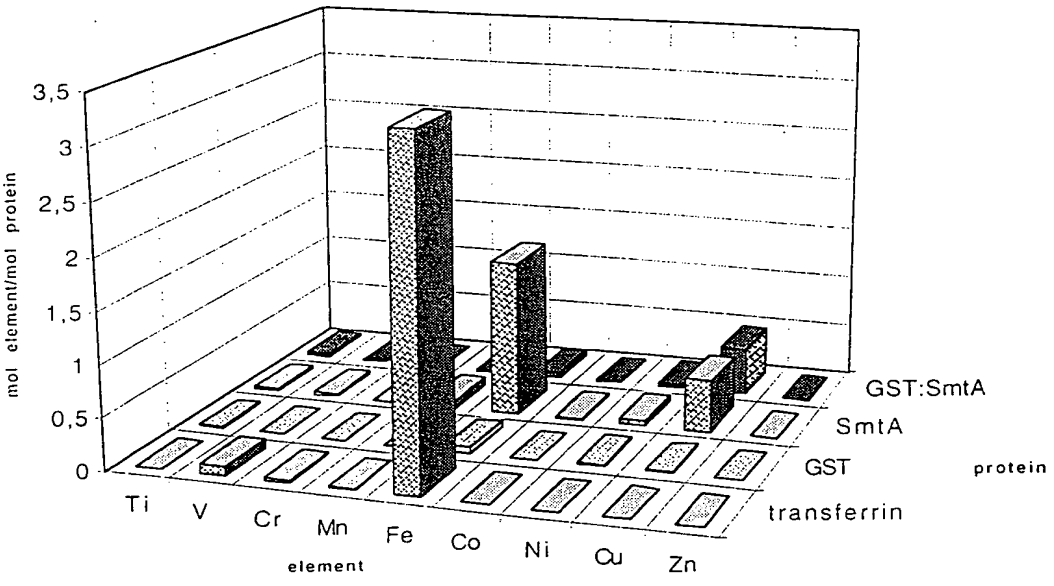
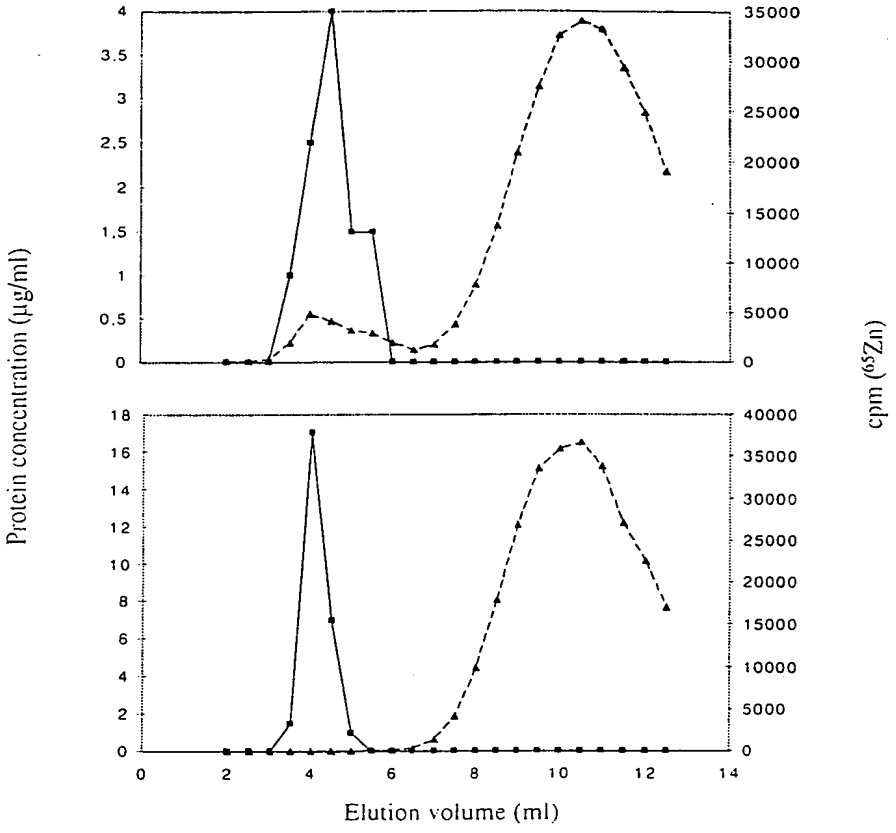
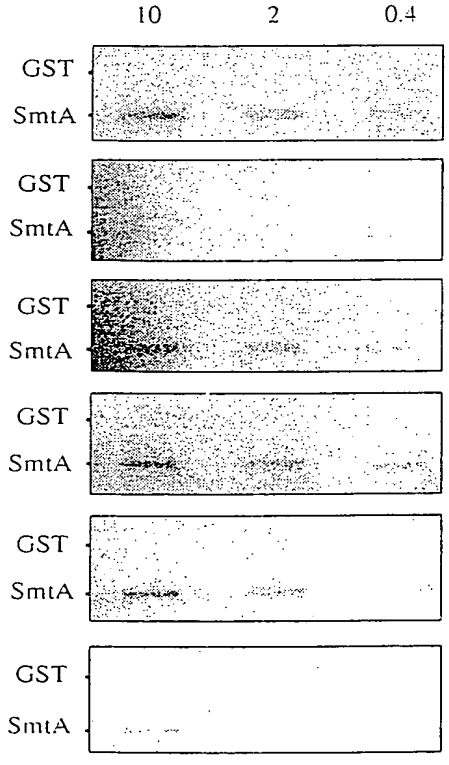

3,5
3
2,5
2
1,5
1
0,5
0
mol element/mol protein
Ti
V
Cr
Mn
Fe
Co
Ni
Cu
Zn
element
GST:SmtA
SmtA
GST
transferrin
protein

FIG. 5

A.



B.



Inhibitor (1 mM)	Percent inhibition
-	-
Zn(II)	88
Fe(III)	0
Fe(II)	0
Mn(II)	12
Cu(II)	74

FIG. 6

PCT/GB99/04045

30/12/99 (P

J.A. Kemp & Co